

What is claimed is:

1. A method of order-ranking document clusters in a plurality of web documents having keywords using entropy data and Bayesian SOM, said method comprising:

5 a first step of recording a query word by a user;

a second step of designing a user profile made up of keywords used for most recent search and frequencies of the keywords, so as to reflect user's preference;

a third step of calculating an entropy value between keywords of each web document and said query word and user profile;

10 a fourth step of collecting data and judging whether data for learning Kohonen neural network is sufficient or not;

a fifth step of ensuring a number of documents using a bootstrap algorithm statistical technique, if it is determined in said fourth step that said data for learning Kohonen neural network is not sufficient;

15 a sixth step of determining prior information to be used as an initial value for each of a network parameter through Bayesian learning, and determining an initial connection weight value of Bayesian SOM neural network model where said Kohonen neural network and Bayesian learning are coupled to one another; and

20 a seventh step of performing real-time document clustering for relevant documents of said plurality of web documents using said entropy value calculated in said third step and Bayesian SOM neural network model.

2. A method according to Claim 1, wherein said seventh step of performing document clustering further comprises the step of calculating entropy value between keywords of each web document and query word given by a user and user profile, and determining a clustering
25 variable.

3. A method according to Claim 1, wherein said prior information determined in advance in said sixth step of determination is in the form of a probability distribution, and said network parameter has a Gaussian distribution.

4. A method according to Claim 1, wherein said number of documents to be ensured
30 by said bootstrap algorithm is fifty.

5. A method according to Claim 1, wherein said document clustering is performed by an average clustering method.

6. A method according to Claim 1, wherein said document clustering is performed by an approach utilizing a distance of statistical similarity or dissimilarity.

7. A method according to Claim 1, wherein said Bayesian SOM is built by K-means method for allocating a relevant document to a nearest document cluster from among a plurality of document clusters disposed around a document.

8. A method according to Claim 7, wherein said K-means method comprises:

a first step of dividing the entire document into K-number of initial document clusters;

a second step of allocating a new document into a document cluster having a centroid which allows shortest distance from each document; and

a third step of repeating said second step of allocating until re-allocation stops,

wherein said K-number of initial document clusters is determined randomly in said step of dividing the entire document,

said centroid of said document cluster receiving said new document has a new value changed from a previous value in said step of allocating a new document, and said repeating step utilizes a seed point if said entire document is divided into random K-number of initial clusters in said step of dividing entire document.